A Ligand That Is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Fluorine Scanning

Bernd Kuhn and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-0446

Received November 29, 1999. Revised Manuscript Received February 4, 2000

Abstract: The associations between biotin and the homologous proteins avidin and streptavidin are two of the most stable complexes found in nature, mainly because of very favorable van der Waals interactions between the ligand and the ideally preformed cavities of the receptors. Previous work, however, suggested that the large binding free energy in these two systems might be further enhanced by small modifications of the valeric acid side chain of biotin. We present in this study a new approach "computational fluorine scanning" as a means to assess how replacing hydrogen atoms with fluorine atoms in ligands could improve ligand binding free energies to their targets. This method allows us to rapidly evaluate and rank the association constants of these two proteins with biotin molecules that are fluoro-substituted on the valeric acid fragment. Our approach is based on a single molecular dynamics simulation of a reference ligand and the postprocessing of the trajectory, in which the reference is replaced by the mutated ligand, using the MM/PBSA free energy analysis. Our results, which are further supported by more elaborate free energy calculations and comparison with experiment, suggest that only one out of 9 investigated fluoro-substitutions, i.e., replacement of the *pro*-8*R* hydrogen atom of biotin with fluorine in the protein avidin, leads to a more favorable binding compared to the naturally occurring complex. Thus, we make a prediction that can be tested experimentally. This computational approach and variations thereof should be useful in the drug design process.

I. Introduction

Two of the most remarkable protein−ligand complexes occurring in nature are the aggregates of D-(+)-biotin (Figure 1) with the protein avidin and its structural homologue streptavidin. Despite the small size of the ligand, biotin forms an unusually strong noncovalent complex with each of these two proteins, their experimental free energies of binding, ΔG_{bind}, being −20.75 and −18.3 kcal/mol for avidin and streptavidin, respectively.[1] This high affinity which is surpassed by only a few noncovalently bonded bioorganic systems, almost exclusively involving metal atoms,[2] has led to many valuable applications in diagnostic assays.[3,4] Elucidating the nature of the binding has motivated a variety of biophysical experiments[5−8] and the X-ray structures of biotin cocrystallized with both avidin[9] and streptavidin[10,11] revealed the protein residues involved in binding. The crystallographic structures show the bicyclo-ring of biotin to be deeply buried into the protein thereby forming several hydrogen bonds and favorable van der Waals (vdW) interactions with almost identical residues in both avidin and streptavidin. Differences in the binding site between the two proteins exist in the coordination of the valeric acid side chain with the CO_2^- terminus being considerably more exposed to solvent in streptavidin than in avidin.

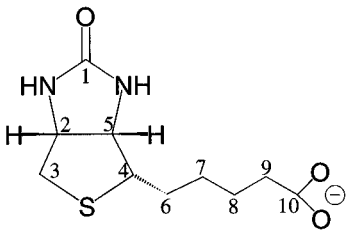

Figure 1. D-Biotin including our numbering of the carbon atoms.The numbers 6−9 denote the CH_2 sites substituted in this study.

Despite the three-dimensional (3D) picture, different opinions existed about the dominant contribution to the high binding affinity, which has motivated several theoretical investigations of these two systems. Miyamoto and Kollman showed in calculations of absolute and relative binding free energies of biotin analogues to streptavidin that the driving force for biotin−protein binding comes predominantly from vdW/nonpolar attractions between the ligand and the preformed cavity of the protein.[12,13] In contrast, electrostatic interaction energies between the ligand and the environment, although considerable in magnitude, are similar in both protein and aqueous solution. This finding was later confirmed by experiment[14] and further

(1) Swamy, M. J. *Biochem. Mol. Biol. Int.* **1995**, *36*, 219−225.
(2) Marchaj, A.; Jacobsen, D. W.; Savon, S. R.; Brown, K. L. *J. Am. Chem. Soc.* **1995**, *117*, 11640−11646.
(3) Wilchek, M.; Bayer, E. A. *Trends Biochem. Sci.* **1989**, *14*, 408−412.
(4) Bayer, E. A.; Wilchek, M. *J. Chromatogr.* **1990**, *510*, 3−11.
(5) Green, N. M. *Adv. Protein Chem.* **1975**, *29*, 85−133.
(6) Torreggiani, A.; Fagnano, C.; Fini, G. *J. Raman Spectrosc.* **1997**, *28*, 23−27.
(7) Torreggiani, A.; Fini, G. *J. Raman Spectrosc.* **1998**, *29*, 229−236.
(8) Torreggiani, A.; Fini, G. *J. Mol. Struct.* **1999**, *481*, 459−463.
(9) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698−710.
(10) Weber, P. C.; Ohlendorf, D. H.; Wendolowski, J. J.; Salemme, F. R. *Science* **1989**, *243*, 85−88.
(11) Weber, P. C.; Wendoloski, J. J.; Pantoliano, M. W.; Salemme, F. R. *J. Am. Chem. Soc.* **1992**, *114*, 3197−3200.
(12) Miyamoto, S.; Kollman, P. A. *Protein-Struct. Funct. Genet.* **1993**, *16*, 226−245.
(13) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci.* **1993**, *90*, 8402−8406.

10.1021/ja994180s CCC: $19.00
Published on Web 04/07/2000

supported computationally by mutating individual nonpolar and polar protein residues that are key fragments for biotin recognition into alanine and calculating the change in free energy of biotin binding.[15]

Given the 3D structure of the binding site and the identification of the dominant binding interactions it is interesting to see if and by which structural modifications one can further enhance the binding affinity between biotin and avidin/streptavidin. Due to the very large binding energy of these two systems this represents a considerable challenge, being reflected in the synthesis of dozens of biotin analogues which all bind less tightly than biotin itself.[5] Using our computational PROFEC[16] analysis we have recently investigated streptavidin bound biotin for possible ligand modifications that could increase the binding affinity.[17] Our results clearly indicate that there is little room for favorable structural changes, especially at the site of the bicyclo-ring, because of the already strong vdW and hydrogen bonding interactions and the high shape complementarity between biotin and the streptavidin cavity. We predicted that changing the *pro*-9*R* hydrogen atom of biotin into the slightly larger CH_3 group could further reduce ΔG_{bind} by 1−3 kcal/mol, depending on the degree of allowed protein flexibility in the calculation. This more favorable binding comes exclusively from the fact that the methyl analogue is less favorably solvated in water compared to biotin. Furthermore, the 9*S*-methyl isomer was predicted to bind less well to streptavidin than its enantiomer 9*R*-methylbiotin. Although our computations of relative solvation and protein−ligand interaction free energies of several biotin derivatives were in many cases in good agreement with experiment, synthesis and binding measurements of the two methyl-substituted compounds found biotin to bind 1 and 2 kcal/mol stronger to streptavidin than the 9*R*- and 9*S*-methylated analogues, respectively.[17]

In this study, we present an alternative computational approach to find a better binding ligand to avidin/streptavidin than biotin. As suggested by the PROFEC analysis, modifications of the $-(CH_2)_4-$ side chain without changing the molecular volume by a large amount are most likely to yield an increase in binding free energy compared to biotin. Due to the considerably greater polarity and polarizability of the C−F bond versus the C−H bond one might expect significant and potentially favorable changes in electrostatic and vdW interactions for some fluorinated biotin analogues. However, calculating the changes in free energy of binding for all eight possible H substitution sites at the valeric acid side chain using a classical free energy perturbation approach would require a considerable amount of computer time.[18,19] This is due to the pairwise comparison intrinsic to standard free energy calculations. As a considerably faster and still effective way to systematically screen the energetic effects of H → F substitutions in biotin we present computational fluorine scanning. This approach is an extension of the method introduced by Srinivasan et al. to study the relative stabilities of different forms of DNA[20] and is closely related to computational alanine scanning, which proved to be powerful in estimating the contribution of individual protein residues to protein−protein interactions.[21] Details of computational fluorine scanning will be given in Section II. To validate the methodology we compare the results of this approach with more elaborate free energy calculations and with experiment in Section III.

II. Methods

1. Model Setup and Equilibration. The starting points for our simulations were the X-ray structures of biotin complexed with egg-white avidin (2.7 Å)[9] and streptavidin (2.6 Å),[10,11] respectively. Both proteins form tetramers and can bind up to four molecules of biotin. Since different binding sites are separated by at least 22 Å in both proteins and are independent from each other we considered only one site in our calculations while keeping the other three as static.

We used the force field representation of AMBER 5 [22] together with new parameters for the biotin ligands, which were taken from a previous study.[12] Biotin and all its analogues were modeled in the anionic state because the valeric acid side chain is likely to be deprotonated at neutral pH. Atomic partial charges of the ligands were calculated using the RESP procedure[23] and are tabulated for biotin in Appendix 1 (Supporting Information). Since the two protein−biotin complexes possess overall charges when assigning default charge states to the residues of the receptor—avidin and streptavidin being highly basic and acidic, respectively—we neutralized the systems by turning off the minimum number of outermost charged residues in each monomer, i.e. residues separated at least 16 Å from the binding site. This charge protocol is based on the assumption that charged residues far away from the active site cancel out their effects on the binding energy and avoids the tedious equilibration of counterions. We completed the system setup by solvating the neutralized biotin−protein complexes with a 20 Å sphere of TIP3P water[24] centered at one of the four biotin molecules.

After energy minimization of the water positions for 1000 steps and molecular dynamics (MD) equilibration of the solvent sphere with fixed solute for 30 ps we minimized the whole system with progressively smaller positional restraints on the solute (from 25 to 0 kcal/[mol·Å^2] for a total of 4000 steps). We subsequently equilibrated our solvated biotin−protein models for 30 ps and, after filling up the 20 Å water sphere with additional TIP3P molecules to compensate for the loss due to diffusion, continued the MD equilibration for an additional 90 ps. The final number of water molecules in each system was 363 and 335 for avidin and streptavidin, respectively. During all steps of the equilibration procedure we used a dual nonbonded cutoff of 12 and 17 Å, the latter being updated only every 20 time steps to account for the slower change in long-range interactions. We only allowed protein residues within 18 Å of one of the biotin binding sites to move, with additional 20 kcal/[mol·Å^2] harmonic constraints for residues between 15.5 and 18 Å from the center of mass of the ligand. The MD simulations were performed at constant $T = 300$ K using the Berendsen coupling scheme,[25] and with a time step of 1.5 fs and the SHAKE algorithm.[26] The solvated and equilibrated biotin−avidin and biotin−streptavidin complexes were used as starting configurations for the following three methods.

2. Fluorine Scanning. Our procedure for doing computational fluorine scanning can be summarized in the following way: (1) Run a single molecular dynamics calculation of one reference ligand bound to the protein in explicit solvent, (2) collect snapshots of the MD

(14) Chilkoti, A.; Tan, P. H.; Stayton, P. S. *Proc. Natl. Acad. Sci.* **1995**, *92*, 1754−1758.
(15) Dixon, R. W.; Kollman, P. A. *Protein-Struct. Funct. Genet.* **1999**, *36*, 471−473.
(16) Radmer, R. J.; Kollman, P. A. *J. Comput.- Aided Mol. Design* **1998**, *12*, 215−227.
(17) Dixon, R. W.; Radmer, R. J.; Kuhn, B.; Kollman, P. A.; Yang, J.; Raposo, C.; Wilcox, C. S.; Klumb L. A.; Stayton P. S. *J. Org. Chem.* Submitted for publication.
(18) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(19) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(20) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(21) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(23) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(25) Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(26) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

trajectory at regular intervals to obtain a representative ensemble of structures, (3) mutate the reference ligand into the desired ligand for each snapshot, and (4) calculate free energy estimates G for the complex, protein, and ligand, respectively, by applying the MM/PBSA energy analysis[20] to the collection of snapshots. Since we treat solvation using a continuum model, the explicit water molecules are discarded from the snapshots before the free energy calculation. Finally, the binding free energy, ΔG_{bind}, can be evaluated as

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \quad (1)$$

As has been shown previously, MM/PBSA is a fast and surprisingly accurate way to calculate free energy differences between protein−ligand complexes,[27] protein−protein complexes,[21] and different forms of DNA and RNA.[20,28] The individual terms of the MM/PBSA approach that contribute to the free energy of a molecule are

$$G_{\text{mol}} = E_{\text{MM}} + G_{\text{solv}} - TS_{\text{solute}} \quad (2)$$

where E_{MM} denotes the sum of intra- and intermolecular mechanical (MM) energies of a molecule in the gas phase, G_{solv} is its solvation free energy, and $-TS_{\text{solute}}$ represents an estimate of the solute entropy. E_{MM} can be further divided into terms arising from electrostatic (E_{es}), van der Waals (E_{vdW}), and internal (E_{int}), i.e. bond, angle, and torsional energies:

$$E_{\text{MM}} = E_{\text{es}} + E_{\text{vdW}} + E_{\text{int}} \quad (3)$$

We calculated this contribution using the *anal* module of AMBER without applying a cutoff for the evaluation of nonbonded interactions. The solvation free energy

$$G_{\text{solv}} = G_{\text{PB}} + G_{\text{np}} \quad (4)$$

consists of an electrostatic term, G_{PB}, which was computed using a finite-difference Poisson−Boltzmann (PB) model and a nonpolar term, G_{np}, which is proportional to the solvent accessible surface area (SA). The PB calculation was done with the *DelPhi* program[29] using PARSE atomic radii[30] and Cornell et al. charges[22] with interior and exterior dielectric constants of 1 and 80, respectively. A grid spacing of 2 grids/Å extending 20% beyond the dimensions of the solute and 750 iterations proved sufficient to get solvation free energies converged to better than 0.001 kcal/mol. G_{np} was calculated from $G_{\text{np}} = \gamma\ \text{SA} + b$ ($\gamma = 0.00542$ kcal/[mol·Å^2], $b = 0.92$ kcal/mol)[30] using the surface area estimation of the program *MSMS*.[31] We estimated the solute entropies with the AMBER module *nmode* from classical statistical formulas and normal-mode analysis of two snapshots which were previously minimized using a distance-dependent dielectric constant to account for solvent screening. Since all ligands investigated differ little from biotin and because of the uncertainty inherent in our estimate of the solute entropy,[20] this was done only once for biotin−avidin and biotin−streptavidin, each. Because of the constant contribution of $-T\Delta S$ for each ligand we quote ΔG^*_{bind}, which is $\Delta G_{\text{bind}} + T\Delta S$, in the following discussion.

We performed fluorine scanning for the avidin system using biotin and 6,6,7,7,8,8,9,9-octafluorobiotin (OFB) as reference ligands, respectively. All eight possibilities of substituting a single hydrogen atom of the $-(CH_2)_4-$ valeric acid fragment with a fluorine atom were considered in our approach and named according to their substitution site. The use of two trajectories originating from different reference ligands allows us to check the consistency of our results. Since collecting snapshots for biotin and postprocessing them for a monofluoro derivative may result in vdW clashes because of the 0.3 Å longer C−F bond length compared to C−H in AMBER, we performed an additional 1000 steps of energy minimization, of the ligand and all protein residues within 5 Å of the ligand, for each mutated snapshot before evaluating the MM/PBSA free energy. In all our calculations we saved a total of 100 snapshots from trajectories of 750 ps, i.e. one structure every 7.5 ps. Mutations of the reference ligand biotin were done by shifting the H atom coordinates of the mutated C−H bond by the ratio of the C−F/C−H bond length along the common bond vector. This procedure was applied accordingly for mutations of OFB to the monofluoro-substituted compounds.

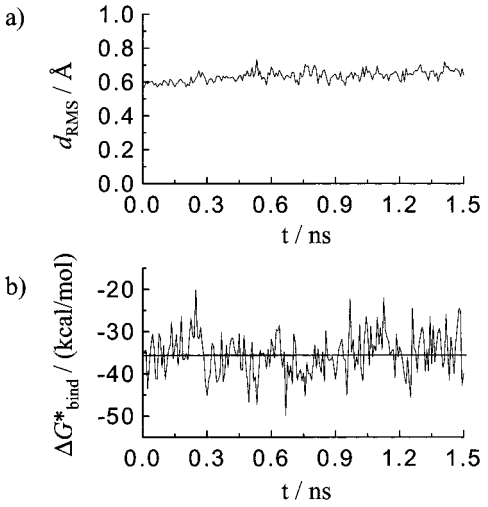

Figure 2. (a) Root-mean-square deviation, d_{RMS}, for the protein backbone atoms of the biotin−avidin complex. (b) Fluctuations of the free energy of binding (ΔG^*_{bind}) between avidin and biotin. d_{RMS} and ΔG^*_{bind} are displayed for the typical 750 ps of data collection (plus an additional 750 ps to show convergence). The straight line in part b indicates the mean of ΔG^*_{bind}.

3. Individual Trajectory MM/PBSA. To assess the error introduced by using the MD snapshots of a different ligand for the MM/PBSA analysis we calculated additional trajectories for two selected monofluoro compounds and compared their MM/PBSA free energies with the results from fluorine scanning. Doing the MD sampling and the postprocessing for the same ligand should provide a more accurate estimate of ΔG_{bind}. Additionally, this was done for the two 9-monomethylbiotin enantiomers, for which experimental data are available.[17] All analogues for which individual MD trajectories were collected were first equilibrated for an additional 90 ps starting from the biotin−avidin and biotin−streptavidin systems, respectively.

4. Free Energy Perturbation. Another way to calculate differences in complexation free energy of two ligands binding to the same protein is to mutate one ligand into another in the presence of both protein and aqueous solution. On the basis of a thermodynamic cycle, $\Delta\Delta G_{\text{bind}}$ can be computed with an accuracy better than 1 kcal/mol in favorable cases.[19] We used the thermodynamic integration (TI) method[18] for perturbing biotin into two monofluorobiotin molecules in both proteins. To calculate the change in solvation free energy we placed the single ligand in a periodic box of ∼3900 water molecules. Initial heating and equilibration of this system for 30 ps resulted in box dimensions of $\sim 50^3$ Å^3. We used the same dual nonbonded cutoff of 12 and 17 Å as for the protein system and performed TI calculations of this system for different simulation times using the *NpT* ensemble ($p = 1$ atm, $T = 300$ K).

III. Results and Discussion

In our comparison of binding free energies of different ligand−protein complexes it is important to ensure that each of these systems is thoroughly equilibrated ahead of data collection. Starting from our solvated and equilibrated biotin−avidin complex, Figure 2a,b plots the root-mean-square deviation, d_{RMS}, of the protein backbone atoms from the X-ray structure and the free energy of binding, ΔG^*_{bind}, along the data collection trajectory. To investigate their convergence, the

(27) Donini, O. T.; Kollman, P. A. *J. Med. Chem.* Submitted for publication.

(28) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265−280.

(29) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.

(30) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(31) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.

Table 1. Energy Contributions to the Free Energy of Binding (ΔG^*_{Bind}) between Avidin and Eight Monofluoro-Substituted Biotin Molecules Using Fluorine Scanning[a]

(a) Reference Trajectory: Biotin

	ΔE_{es}	ΔE_{vdW}	ΔG_{PB}	ΔG_{np}	ΔG^*_{bind}	rank
biotin	−154.4(1.2)	−36.4(0.3)	158.5(0.9)	−3.5(0.1)	**−35.8[b](0.4)**	
6*R*	−153.5(1.1)	−34.3(0.4)	159.8(0.8)	−3.6(0.1)	**−31.6(0.5)**	1
6*S*	−151.9(1.3)	−28.7(0.7)	157.9(0.9)	−3.6(0.1)	**−26.3(0.8)**	8
7*R*	−155.3(1.2)	−32.7(0.5)	161.5(0.9)	−3.6(0.1)	**−30.1(0.6)**	3
7*S*	−151.1(1.3)	−32.3(0.6)	159.3(0.9)	−3.7(0.1)	**−27.8(0.6)**	7
8*R*	−156.2(1.3)	−30.7(0.7)	159.8(0.9)	−3.7(0.1)	**−30.8(0.8)**	2
8*S*	−155.0(1.3)	−28.9(0.9)	159.5(0.9)	−3.6(0.1)	**−28.1(0.9)**	6
9*R*	−148.8(1.2)	−33.6(0.7)	155.8(0.9)	−3.6(0.1)	**−30.3(0.7)**	3
9*S*	−151.2(1.3)	−34.4(0.4)	160.7(1.0)	−3.6(0.1)	**−28.6(0.6)**	5

(b) Reference Trajectory: Biotin. Minimization before MM/PBSA

	ΔE_{es}	ΔE_{vdW}	ΔG_{PB}	ΔG_{np}	ΔG^*_{bind}	rank
biotin	−166.9(1.8)	−37.7(0.2)	160.3(1.2)	−3.1(0.2)	**−47.4(0.6)**	
6*R*	−163.9(1.7)	−38.3(0.2)	160.9(1.2)	−3.4(0.2)	**−44.8(0.6)**	6
6*S*	−165.1(1.8)	−37.3(0.2)	159.6(1.3)	−3.4(0.1)	**−46.2(0.7)**	3
7*R*	−170.5(1.5)	−37.3(0.2)	164.1(1.2)	−3.5(0.2)	**−47.2(0.5)**	2
7*S*	−166.4(1.4)	−37.7(0.2)	162.9(1.1)	−3.5(0.1)	**−44.6(0.5)**	6
8*R*	−166.9(1.7)	−38.2(0.2)	160.8(1.3)	−3.6(0.1)	**−48.0(0.6)**	1
8*S*	−166.5(1.9)	−36.9(0.2)	161.0(1.3)	−3.4(0.2)	**−45.9(0.6)**	4
9*R*	−159.2(1.6)	−38.4(0.2)	157.6(1.2)	−3.5(0.2)	**−43.4(0.6)**	8
9*S*	−160.6(1.6)	−38.7(0.1)	157.9(1.2)	−3.6(0.1)	**−45.0(0.6)**	5

(c) Reference Trajectory: 6,6,7,7,8,8,9,9-Octafluorobiotin (OFB)

	ΔE_{es}	ΔE_{vdW}	ΔG_{PB}	ΔG_{np}	ΔG^*_{bind}	rank
OFB	−131.0(1.2)	−39.6(0.3)	144.5(1.0)	−3.7(0.1)	**−29.8(0.4)**	
biotin	−131.7(1.5)	−35.1(0.4)	144.1(1.3)	−3.8(0.1)	**−26.5(0.5)**	
6*R*	−137.8(1.3)	−36.6(0.3)	154.6(1.1)	−3.7(0.1)	**−23.5(0.4)**	6
6*S*	−138.2(1.3)	−36.5(0.3)	153.5(1.1)	−3.7(0.1)	**−24.9(0.4)**	4
7*R*	−143.6(1.3)	−36.6(0.3)	158.0(1.2)	−3.7(0.1)	**−26.0(0.4)**	2
7*S*	−139.1(1.3)	−36.7(0.3)	155.5(1.1)	−3.7(0.1)	**−23.9(0.4)**	6
8*R*	−142.1(1.3)	−36.5(0.3)	155.6(1.1)	−3.7(0.1)	**−26.7(0.4)**	1
8*S*	−134.5(1.3)	−36.6(0.3)	151.8(1.1)	−3.7(0.1)	**−23.0(0.4)**	8
9*R*	−136.3(1.3)	−36.6(0.3)	151.2(1.1)	−3.7(0.1)	**−25.5(0.4)**	3
9*S*	−130.0(2.2)	−38.3(1.4)	145.8(1.7)	−3.7(0.1)	**−24.5(1.2)**	4

[a] The asterisk indicates that no solute entropy is included. Parts a, b, and c differ in the reference trajectory and in the use of energy minimization before MM/PBSA energy evaluation. The first column shows which H atom of the four valeric acid CH_2 units is replaced by an F atom (see also Figure 1). All energies are averaged over 100 snapshots and are given in kcal/mol. The values in parentheses represent the standard error of the mean. Rank numbers are attributed in 0.5 kcal/mol bins. [b] Adding to ΔG^*_{bind} the change in solute entropy $-T\Delta S \approx 19$ kcal/mol yields $\Delta G_{bind} \approx -17$ kcal/mol.

typical 750 ps of MD sampling were augmented by an additional 750 ps for this system. Both quantities are reasonably converged, with the d_{RMS} slightly rising from 0.60 to 0.65 Å during 1.5 ns. Although ΔG^*_{bind} shows some fluctuations along the trajectory, the mean values of four successive windows of 375 ps each, which are −35.1, −36.5, −35.3, and −34.5 kcal/mol, respectively, indicate little variance.

1. Fluorine Scanning in Avidin. The MM/PBSA analysis allows us to separate the total free energy of binding into electrostatic and vdW solute−solute and solute−solvent interactions, thereby gaining additional insight into the physics of the complex association process. The first row in Table 1a lists these components for the biotin−avidin complex. Clearly, the electrostatic interactions between solute−solute and between solute−solvent change considerably upon dimer formation, however, ΔE_{es} and ΔG_{PB} cancel each other to a great extent. As has been found in previous free energy perturbation[12] and linear interaction energy calculations[32] of this system, it is the increase in vdW energy which is the dominant factor that leads to stable complex formation. In our analysis, the internal energies, E_{int}, cancel each other because we are using a single set of structures for the complex, uncomplexed protein, and uncomplexed ligand. An option would be to additionally compute the free energies for the unbound reactants from the MM/PBSA analysis of separate MD trajectories. Our estimate of the change in solute entropy involved in the binding of biotin to avidin yields 20.1 and 18.2 kcal/mol for two different snapshots of the MD trajectory. Adding the average of these two values to $\Delta G^*_{bind} = -35.8$ kcal/mol results in a free energy of binding of $\Delta G_{bind} \approx -17$ kcal/mol, in good agreement with the experimental value of −20.75 kcal/mol.[1] When comparing absolute binding free energies, it should be kept in mind that our estimate of the solute entropy is approximate due to the use of a normal-mode analysis to calculate the vibrational contribution.

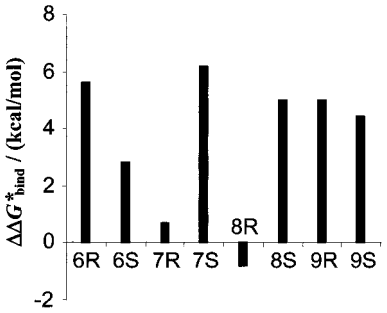

Figure 3. Energy differences obtained in fluorine scanning between the 8 monofluoro-substituted biotin molecules (naming see Table 1) and biotin. The heights of the bars display the sum of the differences using two reference trajectories: biotin (including minimization) and OFB, as listed in Table 1b,c. The asterisk in the change in free energy of binding, $\Delta\Delta G^*_{bind}$, indicates that no solute entropy is included.

Table 1a summarizes the results of our first fluorine scanning in avidin in which we use biotin as a reference ligand. Postprocessing the snapshots of the reference trajectory with the monofluoro analogues yields weaker binding energies for all eight substitutions. While the vdW energies in the complex might be considerably increased due to van der Waals clashes, the electrostatic interactions with their $1/r$ distance dependence should be less sensitive to this artifact and their analysis should still indicate favorable substitution sites. Adding the two contributions $\Delta E_{es} + \Delta G_{PB}$ together we find that just one ligand, 8*R*-fluorobiotin, has more favorable electrostatic interactions with avidin, by 0.5 kcal/mol, than biotin itself. This finding is supported when we minimize the ligand−protein interface before post-processing, as explained in the Methods section. Table 1b shows that with minimization only 8*R*-fluorobiotin (−48.0 kcal/mol) yields a better overall binding free energy to avidin than biotin (−47.4 kcal/mol). The ranking of the ligands is substantially different from that of Table 1a, indicating the presence of significant vdW overlap in some ligand−protein complexes in the first approach. To test the dependence of our predictions on the reference trajectory we list in Table 1c the results for fluorine scanning using the perfluorinated reference ligand, OFB, which is considerably more voluminous than biotin. Because bond lengths shrink for the eight mutated ligands compared to this reference, additional minimization is less important than in the biotin case and was not applied here. Comparing sections b and c in Table 1, which represent our most realistic values for fluorine scanning, we find little differences in the ranking of ΔG^*_{bind}, except for 8*S* and 9*R*. The results of Table 1b,c are summarized in Figure 3 in which we plot the combination of the free energy differences of the monofluoro compounds with respect to biotin using two

(32) Wang, J.; Dixon, R.; Kollman, P. A. *Protein-Struct. Funct. Genet.* **1999**, *34*, 69−81.

Table 2. Energy Contributions to the Free Energy of Binding (ΔG^*_{bind}) between Avidin and Different Biotin Derivatives Using Individual Trajectory MM/PBSA[a]

	ΔE_{es}	ΔE_{vdW}	ΔG_{PB}	ΔG_{np}	ΔG^*_{bind}
biotin	−154.4(1.2)	−36.4(0.3)	158.5(0.9)	−3.5(0.1)	**−35.8[b](0.4)**
8*R*-F	−163.4(1.2)	−36.8(0.5)	167.0(0.9)	−3.5(0.1)	**−36.7(0.5)**
8*S*-F	−135.2(1.3)	−36.7(0.5)	146.5(1.0)	−3.7(0.1)	**−29.1(0.6)**
OFB	−131.0(1.2)	−39.6(0.3)	144.5(1.0)	−3.7(0.1)	**−29.8(0.4)**
9*R*-CH_3	−152.7(1.5)	−39.2(0.4)	160.3(1.0)	−3.5(0.2)	**−35.1(0.6)**
9*S*-CH_3	−131.8(1.1)	−39.5(0.5)	142.6(0.8)	−4.0(0.2)	**−32.6(0.6)**

[a] 8*R*-F, 8*S*-F, 9*R*-CH_3, and 9*S*-CH_3 stand for 8*R*-fluoro-, 8*S*-fluoro-, 9*R*-methyl-, and 9*S*-methylbiotin, respectively. See also the caption footnotes for Table 1. [b] Adding to ΔG^*_{bind} the change in solute entropy $-T\Delta S \approx 19$ kcal/mol yields $\Delta G_{bind} \approx -17$ kcal/mol.

Table 3. Change in Free Energy of Binding, $\Delta\Delta G_{bind}$, for Perturbing Biotin into 8*R*-Fluorobiotin (8*R*-F) and 8*S*-Fluorobiotin (8*S*-F), respectively[a]

	simulation time (ps)			
	37.5	75	150	$\Delta\Delta G_{bind}$
		aqueous solution		
8*R*-F	−16.8/−16.7	−17.4/−	−17.2/−	
8*S*-F	−17.8/−17.5	−17.4/−	−17.1/−	
		avidin		
8*R*-F	−18.1/−18.5	−18.7/−	−18.6/−	**−1.5(0.6)**
8*S*-F	−15.2/−15.6	−15.0/−	−15.9/−	**2.0(0.7)**
		streptavidin		
8*R*-F	−16.2/−16.4	—	−16.8/−	**0.6(0.6)**
8*S*-F	−14.8/−15.0	—	−15.5/−	**2.4(0.7)**

[a] Calculations were performed in aqueous solution, avidin, and streptavidin for different simulation lengths. Numerical values (kcal/mol) are from forward and reverse runs and $\Delta\Delta G_{bind}$ indicates the mean and standard deviation including all simulations.

reference trajectories. The chart illustrates that 8*R* is the only H → F substitution site to increase the binding affinity of biotin. In this analysis, some caution must be applied to the values of 9*R*- and 9*S*-monofluorobiotin. Since inserting an F atom into the biotin or OFB trajectory at these positions results in high intramolecular ligand energies, because of repulsion with the negatively charged CO_2^- terminus, the prerequisite of F-scanning, which is similar configurational space sampling between reference ligand and mutation, is probably least fulfilled here.

2. Other Free Energy Calculations in Avidin. To assess the quality of our fluorine scanning predictions we performed additional free energy calculations for both the 8*R*-F and the 8*S*-F compounds, which are, in the range of ligands, predicted to be the best and one of the weakest binders, respectively. As can be seen from Table 2 this finding is confirmed when we calculate MM/PBSA energies for individual trajectories of the two selected monofluorobiotin molecules. While the 8*R*-F analogue has a 0.9 kcal/mol better binding energy to avidin than biotin, with equal contributions from more favorable vdW and electrostatic interactions upon complex formation, its stereoisomer is predicted to bind 6.7 kcal/mol worse than biotin, exclusively due to an unfavorable change in electrostatic interactions. These values can be further compared with free energy TI calculations for perturbing biotin into 8*R*-F and 8*S*-F, respectively, which are summarized in Table 3. Little hysteresis was found from forward and backward simulations and all numerical values seem to be converged for this small perturbation. As was found with the other methods, replacing an H atom with an F atom at the 8*R* position in biotin improves binding while the corresponding substitution at the 8*S* site worsens ΔG_{bind} to avidin.

Table 4. Changes in Free Energy of Interaction (ΔG_{tot}) between the Ligand and the Protein Residues of Avidin and Streptavidin, Respectively, during the TI Calculation[a]

	ΔG_{es}	ΔG_{vdW}	ΔG_{tot}
	avidin		
Val37B	−0.52	−0.04	−0.56
Trp70B	−0.58	−0.06	−0.64
Arg114B	−1.95	−0.02	−1.97
Trp110D	−0.17	0.75	0.58
Lys111D	−1.38	−0.02	−1.40
	streptavidin		
Trp79B	−0.49	−0.13	−0.62
Arg84B	−0.75	−0.01	−0.76
Lys124D	−0.63	−0.01	−0.64

[a] ΔG_{es} and ΔG_{vdW} are the corresponding electrostatic and vdW components. Only changes with $|\Delta G_{tot}|$ greater than 0.5 kcal/mol are listed.

To gain further insight into the factors leading to the 1−2 kcal/mol improvement in binding energy for 8*R*-fluorobiotin, we analyzed the individual contributions to the free energy change of the 20 protein residues closest to the valeric acid side chain. Table 4 lists those residues whose free energy of interaction with the ligand changes by more than 0.5 kcal/mol during this perturbation. As can be seen from the two dominant contributions, replacing the H atom with the more electronegative F atom considerably improves the electrostatic interactions with both the positively charged Arg114 from the same subunit and Lys111 from a different, dyad-related subunit.

It is interesting to note that in all our three theoretical approaches there is quantitative agreement in the relative binding free energies between biotin and 8*R*-fluorobiotin, while for the 8*S* isomer the individual trajectory MM/PBSA calculation predicts a $\Delta\Delta G_{bind}$ that is ≈5 kcal/mol smaller than in the F-scanning and TI methods. We further investigated the MD trajectories that underlie these methods and found differences in the sampling of the backbone torsional angles of the valeric acid side chain. Figure 4 shows the population of four consecutive torsional angles of the side chain along different trajectories. While the biotin−avidin (Figure 4a) and 8*R*-fluorobiotin−avidin (Figure 4b) complexes sample a similar torsional configuration space, the 8*S*-fluorobiotin−avidin (Figure 4c) system prefers different equilibrium angles for the two torsions $\tau(C_4C_6C_7C_8)$ and $\tau(C_7C_8C_9C_{10})$. Interestingly, as can be seen in Figure 4d, the plot for the TI perturbation from biotin to 8*S*-fluorobiotin resembles the biotin trajectory rather than a mixture of Figure 4a,c, which suggests insufficient sampling in the TI calculation. From analysis of the equilibration of the 8*S*-fluorobiotin−avidin system, which starts from the corresponding equilibrated biotin complex (see Methods section), we find this conformational transition in the side chain to occur at around 30 ps. Hence, it is likely that much longer TI sampling than a few 100 ps would be needed to obtain converged free energies. Because of their similar configurational sampling, it is evident that all three methods presented here yield comparable results for $\Delta\Delta G_{bind}$ between biotin and 8*R*-fluorobiotin. This is not the case for 8*S*-fluorobiotin and consequently leads in F-scanning and TI, due to their basis being the avidin trajectory, to a different binding free energy between this ligand and avidin compared to the individual trajectory MM/PBSA approach. It seems counterintuitive that our MM/PBSA calculation, which uses a more "accurate" torsional sampling for the fluoro compounds than the other two methods, yields a more positive ΔG_{bind}. However, it is the total energy of the ligand−protein complex that determines which ligand conformations are

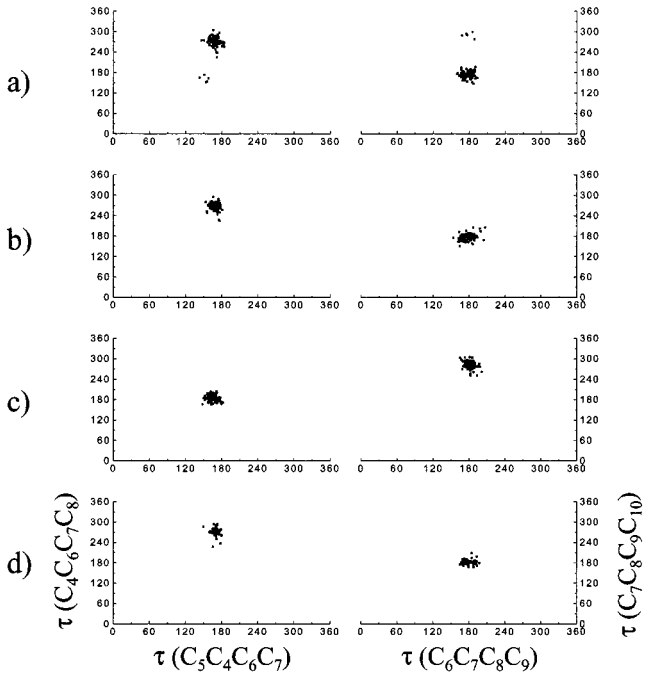

Figure 4. Sampling of four consecutive carbon atom torsional angles, τ, of the valeric acid side chain in biotin during data collection. Parts a, b, and c show τ for the MD trajectory of avidin complexed with biotin, 8*R*-fluorobiotin, and 8*S*-fluorobiotin, respectively (100 snapshots). Part d shows τ during the perturbation from biotin to 8*S*-fluorobiotin in avidin (50 snapshots).

Table 5. Energy Contributions to the Free Energy of Binding (ΔG^*_{bind}) between Streptavidin and Different Biotin Derivatives Using Individual Trajectory MM/PBSA[a]

	ΔE_{es}	ΔE_{vdW}	ΔG_{PB}	ΔG_{np}	ΔG^*_{bind}
biotin	−98.0(1.2)	−36.4(0.4)	116.8(1.1)	−4.0(0.02)	**−21.6[b](0.4)**
8*R*-F	−97.3(1.2)	−34.9(0.4)	114.7(1.1)	−4.0(0.03)	**−21.6(0.4)**
8*S*-F	−84.7(1.2)	−37.7(0.4)	106.0(1.1)	−4.0(0.03)	**−20.4(0.4)**
9*R*-CH_3	−84.1(1.2)	−38.7(0.4)	105.0(1.1)	−4.2(0.05)	**−21.9(0.5)**
9*S*-CH_3	−95.8(1.2)	−37.5(0.4)	117.5(1.1)	−4.1(0.04)	**−19.9(0.4)**

[a] See also footnotes for Tables 1 and 2. Experimental values for $\Delta\Delta G_{bind}$ relative to biotin are +1.3 and +2.2 kcal/mol for 9*R*- and 9*S*-methylbiotin, respectively.[17] [b] Adding to ΔG^*_{bind} the change in solute entropy $-T\Delta S \approx 17$ kcal/mol yields $\Delta G_{bind} \approx -5$ kcal/mol.

accessed in the complex rather than its binding free energy. Our data suggest that binding of 8*S*-fluorobiotin to avidin with the equilibrium torsional angles of the bound biotin (Figure 4a), although having a more favorable ΔG_{bind}, could induce strain in the protein which is relieved by a conformational change of the ligand to the torsional angles depicted in Figure 4c. The absence of the sampling problem for 8*R*-F and the good agreement between all three theoretical methods is strong support that this compound could form a more stable complex with avidin than biotin.

3. Streptavidin. Only little differences exist in the biotin binding site between avidin and its homologue streptavidin, almost exclusively around the valeric acid side chain region, which makes it interesting to compare the stereoselectivity for 8-fluoro substitutions in both proteins. Results for the free energies of binding of biotin and 8*R*- and 8*S*-fluorobiotin in streptavidin are listed in Tables 3 and 5 for the TI and individual MM/PBSA methods, respectively. Both computational approaches agree well in their estimates of $\Delta\Delta G_{bind}$ and predict, in contrast to avidin, that 8*R*-fluorobiotin should bind as well as biotin. An analysis of the dominant TI contributions to the free energy change in the protein is given in Table 4.

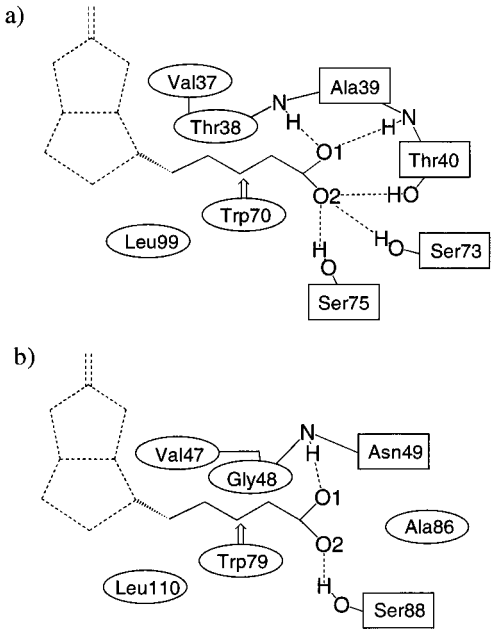

Figure 5. Schematic comparison of the binding situation around the valeric acid side chain in (a) avidin and (b) streptavidin, as given by the X-ray structures. Rectangles and ellipsoids represent residues involved in hydrogen bonding and vdW interactions, respectively. An arrow indicates that this residue is located below the plane of the paper. No solvating water molecules are shown.

Adding the solute entropy correction of ≈17 kcal/mol to the absolute free energies of binding of Table 5 we obtain $\Delta G_{bind} \approx -5$ kcal/mol for the biotin−streptavidin complex. The relatively large difference to the experimental value of −18.3 kcal/mol[1] is surprising given the good agreement between theory (−17 kcal/mol) and experiment (−21 kcal/mol) in the absolute binding free energy between biotin and the homologous protein avidin. Replacing our RESP fitted charges with PARSE charges, which were derived together with the PARSE radii to reproduce small molecule hydration free energies,[30] we find only a slight improvement of 1 kcal/mol in the computed $\Delta\Delta G_{bind}$ between the two proteins. This rules out that the use of AMBER charges in our approach, which is needed to be consistent with the MM calculation, is at the origin of this free energy discrepancy.

To investigate the difference in absolute free energies further we schematically illustrate in Figure 5 the structural dissimilarities in the biotin binding site between the two proteins, as revealed by the X-ray structures of the biotin−protein complexes. The residues forming the nearest neighbor interactions to the bicyclo-ring are not shown because they are identical in the two proteins, with the exception that Phe79 in avidin is replaced by Trp92 in streptavidin. The major difference between the two binding sites consists of the degree of solvent exposure of the valeric acid side chain. While the CO_2^- terminus of biotin forms five hydrogen bonds to protein residues in avidin these interactions are reduced to two in streptavidin, the others being replaced by interactions with solvent water molecules. Our calculations reveal that upon complexation the solvent accessible surface area of the CO_2^- group of biotin reduces on average in avidin from 37 Å^2 to 4 Å^2 and in streptavidin from 36 Å^2 to 29 Å^2. This large difference in solvation contact between the two biotin−protein complexes is also reflected in the PBSA energy contributions of Tables 2 and 5, which show in avidin a 42.2 kcal/mol higher desolvation penalty ($\Delta G_{PB} + \Delta G_{np}$) upon complexation than in streptavidin.

While the larger free energy price for desolvation in avidin is compensated by more favorable MM interaction energies

between the ligand and the protein in the complex, it is clear that the gas-phase interaction energies, as calculated by AMBER, and the continuum model PB energies have to be well 'balanced', i.e. errors in relative MM interaction and continuum solvation energies have to be similar, to yield accurate absolute binding free energies. One apparent problem is that hydrogen bonding and other first-solvation-shell effects in continuum solvent descriptions[33] are accounted for only in an average way through their parametrization on small molecule solvation energies while molecular mechanics treats hydrogen bonds explicitly. Since the hydration free energy of a CO_2^- group is large (e.g. CH_3COO^-: experimental $\Delta G_{solv} = -80.7$ kcal/mol)[30] and the continuum model parameters were optimized only for fully hydrated molecules we would expect a significant absolute error in ΔG_{PB} for the partially solvated carboxyl group in the biotin−streptavidin complex.

To test the hypothesis that an inadequate PB treatment of first-solvation-shell effects for the partially hydrated biotin is at the origin of the differences between the calculated binding free energies of the two biotin−protein complexes, we recomputed ΔG_{bind} with the MM/PBSA method for the streptavidin system, but now explicitly including two of the water molecules which are hydrogen bonded to the CO_2^- group of biotin as part of the complex. This choice of two water molecules was due to the fact that in the MD trajectories, the first solvation shell of the biotin CO_2^- terminus in streptavidin has on average 2.2 more water molecules than in avidin. Although this is a crude approach, it allows us to treat the complex formation between biotin and avidin or streptavidin with an equal number of explicit hydrogen bonds. Using TIP3P charges and TIP3P charges that are scaled by 70% to reproduce the gas-phase dipole moment of water, we obtain for the biotin−streptavidin complex $\Delta G^*_{bind} = -43.0$ and -30.4 kcal/mol, respectively. Taking into account the change in solute entropy in the biotin−streptavidin complex of $-T\Delta S_{solute} \approx +17$ kcal/mol and using the model with two "gas phase" like water molecules we find a very similar error of 4−5 kcal/mol in ΔG_{bind} compared to experiment for both avidin and streptavidin. Although this correction is rough, it supports the assumption that an inadequate treatment of first-solvation-shell effects of the CO_2^- group of biotin is the reason for the inability of MM/PBSA to accurately reproduce the absolute binding free energy of biotin−streptavidin. It should be noted that our comparisons of relative ΔG^*_{bind} values of similar ligands to the same protein in this study are much less prone to this error because of their very similar solvation properties.

To reduce the error in the absolute MM/PBSA energies, one could consistently include the first solvation shell as explicit water molecules into the MM treatment. However, this would require more time-consuming, individual MD trajectories for the complex as well as the separated protein and ligand, each. Alternatively, a refinement of the dielectric radii for the PB treatment in a way that ($\Delta G_{PB} + \Delta G_{np}$) reproduces both experimental hydration free energies of a test set of molecules and free energy data of partially solvated functional groups, which could come from high-level ab initio calculations, might increase the accuracy of our MM/PBSA approach. Presumably, this would require a more diverse set of dielectric radii than currently used and a surface area dependent H-bonding correction term, as was used by Marten et al.[34] The question of more appropriate atomic radii in the MM/PBSA treatment will be addressed in detail in a later publication.

One might argue that our approximation of using the snapshots of the complex trajectory as the basis for the conformations of the uncomplexed protein and ligand is differently fulfilled for avidin and streptavidin, thereby contributing to the too large computed difference $\Delta\Delta G_{bind}$. While a quantitative assessment of this error would require the generation of individual trajectories for the two proteins and biotin itself, this study finds that the MM/PBSA calculated G_{biotin} differs only by 1 kcal/mol between the avidin and the streptavidin simulations. This small difference indicates that the biotin snapshots generated in the complex with avidin and streptavidin, respectively, yield similar free energies and hence that the computed $\Delta\Delta G_{bind}$ would not profit from individual ligand trajectories.

4. 9-Methylbiotin Isomers. Using our individual trajectory MM/PBSA approach, we also calculated free energies of binding for the two 9-methylbiotin isomers, for which experimental data are available.[17] In streptavidin, the experiment found a 1.3 and 2.2 kcal/mol preference in ΔG_{bind} for biotin compared to 9*R*- and 9*S*-methylbiotin, respectively. As is listed in Table 5, MM/PBSA yields free energy differences of -0.3 and 1.7 kcal/mol, respectively. Additionally including the change in solute entropies upon complexation one would expect the calculated relative free energies of binding between the methylated compounds and biotin to become slightly more positive which would further improve our agreement with experiment. This is because of the larger number of degrees of freedom in methylbiotin that become restricted upon complexation. However, because this energy correction is likely to be smaller than the uncertainty of our normal-mode analysis, which shows fluctuations of ±2 kcal/mol for different snapshots, we did not include the entropy correction here.

In the avidin case, only calorimetric measurements for the 9-methyl-substituted ligands exist which resulted in a $\Delta\Delta H_{bind}$ in favor of biotin between 5 and 9 kcal/mol, depending on the pH, and which predicted that the 9*R* isomer binds by ∼1 kcal/mol better than the 9S compound. While experimental ΔG_{bind} values would be needed for a quantitative comparison with our calculation, Table 2 shows that we reproduce the correct order in binding energy among biotin and the two isomeric 9-methylbiotin derivatives.

IV. Summary and Conclusions

In this study, we have investigated the interaction free energies between avidin and streptavidin complexed with several biotin analogues to (a) assess the quality of computational fluorine scanning as a means to rapidly and efficiently rank ΔG_{bind} of a set of ligands and (b) propose mutations of biotin which will increase its binding affinity to one of the proteins. Our results indicate that F-scanning is able to reproduce relative binding free energies at a much lower computational cost than more elaborate free energy calculations. The time economy of the scanning approach to evaluate relative energies compared to a conventional free energy perturbation calculation is 2-fold. First, we do not have to sample intermediate points of a perturbation, but just one endpoint. Second, a set of *N* ligands can be ranked by performing *N* fast MM/PBSA analyses instead of the pairwise comparison of *N* time-consuming MD trajectories. Clearly, this approach is based on the assumption that the mutated ligands and the reference ligand sample a similar configurational space in their dynamics, and hence is likely to be most useful for small mutations, as was the purpose of this

(33) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(34) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−11788.

work. Future studies on more diverse ligands will more specifically reveal the limitations of this approach.

A computationally more expensive, but very promising method to compute ΔG_{bind} consists of calculating the MD structures for each ligand individually and to postprocess them with MM/PBSA. We found very good correlation with experimental relative free energies for the 9-methylbiotin compounds, and calculations on a set of diverse avidin ligands, with binding free energies ranging between $\Delta G_{bind} = -4.5$ and -20.4 kcal/mol, are in excellent quantitative agreement with experiment.[35] It was shown in this study that typical TI simulation lengths of a few hundred picoseconds, although appearing to be converged, are not long enough to accurately sample the change in torsional conformation of the valeric acid fragment when perturbing biotin into 8*S*-fluorobiotin in the protein−ligand complex. Because of the more extensive sampling of the endpoints, individual trajectory MM/PBSA is able to capture the different conformational preferences and the energetic effects involved. It is clear, as was also stated previously,[36,37] that much longer TI simulation times than generally applied or enhanced sampling techniques[38,39] should be used to obtain convergence.

We found that individual trajectory MM/PBSA led to an excellent agreement for the absolute free energy of association of the biotin−avidin system ($\Delta\Delta G_{bind} = 4$ kcal/mol), but the homologous biotin−streptavidin complex had an absolute ΔG_{bind} too small in magnitude by 13 kcal/mol. Our analysis suggests that this is due to an insufficient description of the first-solvation-shell effects of the partially solvated CO_2^- group of biotin in the complex with streptavidin using a continuum method. Errors in absolute free energies are relatively large because of the considerable solvation free energy of this negatively charged functional group. However, in this study, in which we compare relative binding free energies of very similar ligands to the same protein, this error is unlikely to have a major effect. The lack of continuum model parameters that are refined for partial hydration and our results for the biotin−streptavidin complex suggest that in cases where charged functional groups are only partly buried in the complex, one might expect an error in the MM/PBSA calculation of ΔG_{bind}. To further increase the accuracy of our MM/PBSA approach we are working on a more refined continuum model parameter set, which also includes free energies of partially solvated molecules, and on correction terms for better incorporating first-solvation-shell effects.

On the basis of the accuracy of the individual MM/PBSA calculation in reproducing experiment and the quantitative agreement between all three theoretical methods we propose that replacing the *pro*-8*R* hydrogen with a fluorine atom should increase the binding affinity to avidin by 1−2 kcal/mol. This is partly due to better vdW interactions with the protein binding site and due to favorable electrostatic interactions with the residues Arg114 of the same subunit and Lys111 of a different, dyad-related subunit. The synthesis of this compound and measurement of its free energy of binding to both proteins is envisaged.

More generally, our MM/PBSA method gives us an independent way to calculate protein−ligand binding free energies that can complement standard free energy calculations. In the case of 9*R*-methylbiotin binding to streptavidin, both methods calculate the free energy of binding relative to biotin as too favorable compared to experiment, albeit the MM/PBSA value is significantly closer to the correct value. In the case of 8*R*-fluorobiotin binding to avidin, the two methods lead to a $\Delta\Delta G_{bind}$ relative to biotin of -0.9 and -1.5 kcal/mol, which lends more credibility to this prediction.

Acknowledgment. We would like to thank O. Donini for helpful discussions. B.K. gratefully acknowledges support from the German Academic Exchange Service (DAAD) through a research scholarship. P.A.K. is pleased to acknowledge research support through the NIH (GM-29072), supercomputer support from the NSF supercomputer centers, and graphics support from the UCSF computer graphics lab, T. Ferrin, P.I. (RR-1081).

Supporting Information Available: Appendix 1 giving atomic partial charges for the biotin molecule (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA994180S

(35) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* Submitted for publication.
(36) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532−4545.
(37) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487−1493.
(38) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(39) Simmerling, C.; Fox, T.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 5771−5782.